OM protein - protein search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on:    December 27, 2001, 16:54:43 ; Search time 28.15 Seconds
                                         (without alignments)
                                         633.209 Million cell updates/sec

Title:          US-09-830-647-2
Perfect score:  1206
Sequence:       1 MNSGAMRIHSKGHFQGGIQV..........LKPPVKVEDMSQSPAVHLM 234

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters:    219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    PIR_68:*
              1:   pir1:*
              2:   pir2:*
              3:   pir3:*
              4:   pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Match Length | DB | DB ID | Description |
|---|---|---|---|---|---|---|
| 1 | 784 | 65.0 | 601 | 2 | T02633 | hypothetical prote |
| 2 | 101.5 | 8.4 | 779 | 2 | T01304 | hypothetical prote |
| 3 | 100.5 | 8.3 | 615 | 2 | S38008 | hypothetical prote |
| 4 | 94 | 7.8 | 1415 | 2 | S52267 | DNA polymerase III |
| 5 | 91 | 7.5 | 908 | 2 | S51293 | probable membrane |
| 6 | 91 | 7.5 | 994 | 1 | A47474 | NAD+ ADP-ribosyltr |
| 7 | 90 | 7.5 | 219 | 2 | C71927 | cag island protein |
| 8 | 89 | 7.4 | 710 | 2 | S26006 | hypothetical prote |
| 9 | 89 | 7.4 | 871 | 2 | T48502 | hypothetical prote |
| 10 | 89 | 7.4 | 1119 | 2 | B70126 | surface-located me |
| 11 | 87.5 | 7.3 | 392 | 2 | B70242 | conserved hypothet |
| 12 | 87.5 | 7.3 | 506 | 2 | T28810 | hypothetical prote |
| 13 | 87.5 | 7.3 | 520 | 2 | A71564 | hypothetical prote |
| 14 | 87.5 | 7.3 | 885 | 2 | B69783 | transporter homolo |
| 15 | 87 | 7.2 | 426 | 2 | F82315 | conserved hypothet |
| 16 | 86.5 | 7.2 | 690 | 2 | T34149 | hypothetical prote |
| 17 | 86.5 | 7.2 | 1274 | 2 | T02636 | DI protein homolog |
| 18 | 86.5 | 7.1 | 1462 | 2 | T06819 | DNA topoisomerase |
| 19 | 86 | 7.1 | 1081 | 2 | T15692 | hypothetical prote |
| 20 | 85 | 7.0 | 273 | 2 | E81330 | probable periplasm |
| 21 | 85 | 7.0 | 602 | 2 | D69331 | probable DNA topoi |
| 22 | 85 | 7.0 | 773 | 2 | T00502 | probable receptor- |
| 23 | 85 | 7.0 | 299 | 2 | S34588 | senescence marker |
| 24 | 84.5 | 7.0 | 942 | 2 | JC7316 | testicular zinc fi |
| 25 | 84.5 | 7.0 | 1891 | 2 | T13594 | hypothetical prote |
| 26 | 84.5 | 7.0 | 1920 | 2 | T18893 | gene hindsight pro |
| 27 | 84 | 7.0 | 324 | 2 | G81330 | probable phosphata |
| 28 | 84 | 7.0 | 1059 | 1 | A35210 | ferroxidase (EC 1. |
| 29 | 83.5 | 6.9 | 362 | 2 | T48564 | probable serine ri |
| 30 | 83.5 | 6.9 | 833 | 2 | H72205 | maltose ABC transp |
| 31 | 83.5 | 6.9 | 1079 | 2 | T18356 | membrane protein p |
| 32 | 83.5 | 6.9 | 5105 | 2 | T32650 | hypothetical prote |
| 33 | 83 | 6.9 | 650 | 2 | A55551 | dnaK-type molecula |
| 34 | 83 | 6.9 | 1176 | 2 | S66771 | probable DNA-direc |
| 35 | 83 | 6.9 | 1319 | 2 | T28203 | hypothetical prote |
| 36 | 83 | 6.9 | 1854 | 2 | T13576 | hypothetical prote |
| 37 | 82.5 | 6.9 | 244 | 2 | T25531 | hypothetical prote |
| 38 | 82.5 | 6.8 | 430 | 2 | T25623 | hypothetical prote |
| 39 | 82.5 | 6.8 | 434 | 2 | G64444 | amidase - Methanoc |
| 40 | 82.5 | 6.8 | 482 | 2 | A30198 | dihydrolipoamide S |
| 41 | 82.5 | 6.8 | 592 | 1 | LLBY | actin-binding prot |
| 42 | 82.5 | 6.8 | 946 | 1 | A71843 | d-lactate dehydrog |
| 43 | 82.5 | 6.8 | 1225 | 2 | A49464 | chromosome segrega |
| 44 | 82 | 6.8 | 297 | 1 | MNVNRV | nonstructural prot |
| 45 | 82 | 6.8 | 304 | 2 | JX0209 | lectin, galactose/ |

ALIGNMENTS

RESULT 1
T02633
hypothetical protein RG135C18.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C;Accession: T02633
R;Kellen, J.; Burkhart, J.
A;Reference number: Z14683
A;Description: The sequence of Homo sapiens BAC clone RG135C18.
A;Molecule type: DNA
A;Residues: 1-601 <KEL>
A;Cross-references: EMBL:AC005164; NID:g3242749; PIDN:AAC23786.1; PID:g3242750
A;Genetics:
A;Map position:
A;Introns: 60/3; 77/3; 101/1; 126/3; 139/1; 154/2; 197/2; 235/3; 277/2
A;Note: WUGSC:H_RG135C18.1

Query Match     65.0%;  Score 784;  DB 2;  Length 601;
Best Local Similarity  100.0%;  Pred. No. 3.1e-55;
Matches  154;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     74 RVEEFLSKDISYLISNKKEARKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDT 133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 RVEEFLSKDISYLISNKKEARKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDT 60

Qy    134 VCLSRGKLLVEKAIRKDHDFIPSNSILSNALSWGVKILHIDDIRYYIEQKKKELYLLKKSS 193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VCLSRGKLLVEKAIRKDHDFIPSNSILSNALSWGVKILHIDDIRYYIEQKKKELYLLKKSS 120

Qy    194 TSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQ 227
          ||||||||||||||||||||||||||||||||||
Db    121 TSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQ 154

RESULT 2
T01304
hypothetical protein T14P8.6 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 18-Feb-2000
C;Accession: T01304
R;Kalicki, J.; Elliott, G.; Cloud, J.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of A. thaliana T14P8.
A;Reference number: Z14290
A;Accession: T01304
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-779 <KAL>

```
A;Cross-references: EMBL:AF069298; NID:g3193282; PID:g3193291
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Introns: 94/3; 202/3; 254/3; 562/3; 585/3; 610/2; 632/2; 667/1; 727/3
A;Note: T14P8.6
C;Superfamily: DNA mismatch repair protein

Query Match     8.4%;    Score 101.5;    DB 2;    Length 779;
Best Local Similarity  22.9%;    Pred. No. 2.7;
Matches   48; Conservative  35; Mismatches  66;  Indels  61; Gaps  9;

Qy    23  EKNRPSLKSLKTDNRP-----EKSK------CKPLWGK--------VFYLD-------LPSVTISE  62
              ||:||  |||||||||          ||   |:|             | :||          | :||
Db   392  EKENPSLREVEIDNSSPMEKFKFEIKACGTKKGEGSLSVHDVTHLDKTPSKGLPQLNVTE  451

Qy    63  KLQDKIDLGGRVEEFLSKDISYLISNKKEAKFAQ--------TLGR------ISPVPSPES  110
              ||:|| |     |  ||:|| | |||  |||  |         |  |:       |:|| |||
Db   452  KVTDASKDLSSR----------------SSFAQSTLNTFVIMGKKHENISTILSEIP  493

Qy   111  AYTAETTSPHPSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPS--NSILSNALSWGVK  168
              |  |:  | |  |||  ||| |||| |||||  |  |||||||    ||:|||||||| |
Db   494  VLRNQTSSYRVEKSKFEVRALASRCLVEGDQLDDMVISKEDMTPSERDSELGNRISPGTQ  553

Qy   169  ILHID--------DIRYIEQKKKELYLL  189
              ::|           ::|:||:|:|:::
Db   554  ADNVERHERVLGQFNLGFIIAKLERDLFIV  583
```

RESULT  3
```
S38088
hypothetical protein YKR019c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Oct-1999
C;Accession: S38088
R;Duesterhoeft, A.; Moestl, D.; Poehlmann, R.; Philippsen, P.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37811
A;Accession: S38088
A;Molecule type: DNA
A;Residues: 1-615 <DUE>
A;Experimental source: strain S288C
A;Cross-references: EMBL:Z28244; NID:g486438; PID:g486439; GSPDB:GN00011; MIPS:YKR019c
C;Genetics:
A;Gene: MIPS:YKR019c
A;Map position: 11R

Query Match     8.3%;    Score 100.5;    DB 2;    Length 615;
Best Local Similarity  20.8%;    Pred. No. 2.4;    Mismatches  121;  Indels  67; Gaps  9;
Matches   60; Conservative  41;

Qy     3  SGAMRIHSKGHFPQGGIQVKNEKNRPS-----------LKSLKTDNRPEKSKCKPLWGKV  50
              |:|  |||:| ||           |||             |||||||||||||||||||
Db   150  NGLLGIHSRSSSRNG---SNESLTPGQRTPDNRSQENLTSFSSGRRLSSSSMEPATNKD  206

Qy    51  FYLDLPSVTISEKLQKDIKDLGGRVE-------------EFLSKDISYLISNKKEAKFAQT  98
              |||| |    ||:||:   |||:   |            :|||:||||| |||||||||:|
Db   207  SNKALPKRRPSPPLQSSLVGSGQLHENENLSSISIDSRHSLNPDTSDVISNRSQTSLSQT  266

Qy    99  LGRISPVPSPESAYTAETTSPHPSHDGSSFKS--PDTVCLSRGKLLVEK-----------  145
              |||:|||||||||           ||  ||    |||| |||||||||
Db   267  INQSLCESEPSIASSNTTT-TSNQGSGLPNLVPNYSSDMRKKLVNKFKRKVFGSKPK  325

Qy   146  ------------AIKDHDFIPS--------NSILSNALSWGVKILHIDDIRYIEQKK  183
                          |:||:|| ::        ||:||:||||||||||||||:||:||
Db   326  HLSSQYEMDASSEELGQHEGQPSMRFKTTLRKTSVSTNAENDHASSLHEGNLRYKY-NPS  384

Qy   184  KELYLLLKKSSTSVRDGKRVSGAQKTRGCRLKKPFVKVEDMSQSPAVH  232
              |  | ||||||||||||  ||||||
Db   385  NDTDVDYDDYDDSESDQNQDALTKPRKRDRIKR---KIRNSANKTAHH  430
```

RESULT  4
```
S52267
DNA polymerase III - Staphylococcus aureus
C;Species: Staphylococcus aureus
A;Residues: 1-1415
C;Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 24-Sep-1999
A;Accession: S52267
A;Reference number: S52267
A;Description: Staphylococcus aureus DNA polymerase III.
A;Molecule type: DNA
A;Status: Preliminary
A;Cross-references: EMBL:Z48003; NID:g642269; PIDN:CAA88043.1; PID:g642270
C;Superfamily: DNA-directed DNA polymerase III alpha chain polC

Query Match     7.8%;    Score 94;    DB 2;    Length 1415;
Best Local Similarity  24.0%;    Pred. No. 23;
Matches   63; Conservative  34; Mismatches  93;  Indels  72; Gaps  14;

Qy    18  IQVKNEKNRPSIKSLKTDNRP-EKSKCKPLWGKWFYLDLPSVTISEKLQKDIKDLGGRVE  76
              |:| |||||:||||||||||  |  | |||| |  |  |  | ||:|| ||:||||:||
Db   805  VPIRDELYTPRMEGANEEIRELSYANARKLYGE----DLPQIVI-DRLEKELKSIIGN-  857

Qy    77  EFLSKDISYLISN--RKEAKFAQTLGRISPVPSPESAYTAETT--SPHPSH-------  123
              |:      |  ||    | || | | | | || |  | | ||
Db   858  ---GFAVIYLISQRLVKKSLDDGIVLGSRGSVGSSFVATMIEITEVNPLPPHYICPNCKT  914

Qy   124  -------DGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDI  175
                     |||:|||||||||||||||||||:|||||||||:|||||||||||||||||
Db   915  SEFFNDGSVGSGFDLPDKTCETCGAPL----IKEGQDIPFEKFLGFK---GDKVPDI-DL  966

Qy   176  RYYIEQKKKELYLLLKKSSTSVRDGGKRV---------GSGAQKTRG---  213
              | ||:|||:|| |   | |   |  |  |          |:||  |
Db   967  NFSGEYQPN----AHNYTKVLFGEDKVFRAGTIGTVAEKTAFGYVKGYLNDQGIHKRGA  1021

Qy   214  --RLKKPFVKVEDMS--QSPA  230
                 | | |  | ||:   |:|
Db  1022  EIDRLVKGYMDIYDFTPIQYPA  1043
```

RESULT  5
```
S51293
probable membrane protein YNL321w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N0339
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 23-Mar-2001
C;Accession: S51293; S59562; S63302
R;Nicaud, J.J.
submitted to the EMBL Data Library, January 1995
A;Description: Sequence analysis of a 13.9 kb fragment of yeast chromosome XIV identi
A;Reference number: S51293
A;Accession: S51285
A;Molecule type: DNA
A;Residues: 1-908 <NIC>
A;Cross-references: EMBL:Z46259; NID:g633655; PID:g633664
R;Mafahi, M.; Nicaud, J.M.; Levesque, H.; Gaillardin, C.
Yeast 11, 1077-1085, 1995
A;Title: Sequencing analysis of a 24.7 kb fragment of yeast chromosome XIV identifies
A;Accession: S59562; MUID:96076632
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-908 <MAF>
A;Cross-references: EMBL:Z46259; NID:g633655; PIDN:CAA86376.1; PID:g633664
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
R;Mattahi, M.; Nicaud, J.M.; Levesque, H.; Gaillardin, C.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63287
A;Accession: S63302
```

A;Molecule type: DNA
A;Residues: 1-908 <MAW>
A;Cross-references: EMBL:Z71597; NID:g1302433; PID:e239767; PID:g1302434; MIPS:YNL321w
A;Experimental source: strain S288C
A;Genetics:
A;Map position: 14L
C;Keywords: transmembrane protein
F;250-266/Domain: transmembrane #status predicted <TM1>
F;413-429/Domain: transmembrane #status predicted <TM2>
F;496-512/Domain: transmembrane #status predicted <TM3>
F;531-547/Domain: transmembrane #status predicted <TM4>
F;561-577/Domain: transmembrane #status predicted <TM5>
F;595-611/Domain: transmembrane #status predicted <TM6>
F;631-647/Domain: transmembrane #status predicted <TM7>
F;689-705/Domain: transmembrane #status predicted <TM8>
F;748-764/Domain: transmembrane #status predicted <TM9>
F;820-836/Domain: transmembrane #status predicted <TM10>
F;886-902/Domain: transmembrane #status predicted <TM11>

```
Query Match             7.5%;  Score 91;  DB 2;  Length 908;
Best Local Similarity   31.1%;
Matches    41;  Conservative    19;  Mismatches   42;  Indels   30;  Gaps    9;

Qy   32 LKTDNRPEKSKCKP--------LW-GKVFYLDLPSVTISEKLQKDIKIDLGGRVEEFLSKDI 83
           |:| ::|| :| :| :| |:|  |||||| : | ||| || :
Db  271 LVTNSAKEYSKCLYKLANYFLMPFGKMVYL---LQDEQYLQED-KDEGISMQQFYNWVT 325

Qy   84 SY---LISNKKEAKFAQTLGRISPVPSPESAY--TAETT-----SPHPSHDGSSFKSPD 132
        ||   ||:: |:|||||:|:|||||||:||||  |||::     |||||||||||||||
Db  326 SYSNRLVFHQSQAKFQOR--EDHPAPATESSSLMPPANTTATPLNSNHPSYNSIRHEIPH 383

Qy  133 TVC---LSRGK 140
            |   ||||
Db  384 AAAQRRYFGRGK 395
```

RESULT 6
A47474
NAD+ ADP-ribosyltransferase (EC 2.4.2.30) - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: A47474
R;Uchida, K.; Hanai, S.; Ishikawa, K.; Ozawa, Y.; Uchida, M.; Sugimura, T.; Miwa, M.
Proc. Natl. Acad. Sci. U.S.A. 90, 3481-3485, 1993
A;Title: Cloning of cDNA encoding Drosophila poly(ADP-ribose) polymerase: leucine zipper
A;Reference number: A47474; MUID:93234521
A;Accession: A47474
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-994 <UCH>
A;Cross-references: GB:D13806; GB:D13807; GB:D13808; NID:g303545; PIDN:BAA02964.1; PID:g
A;Note: sequence extracted from NCBI backbone (NCBIN:129703, NCBIP:129704)
C;Superfamily: NAD+ ADP-ribosyltransferase
C;Keywords: DNA binding; glycosyltransferase; NAD; nucleus; pentosyltransferase
C;Genetics:
A;Gene: Parp

```
Query Match             7.5%;  Score 91;  DB 1;  Length 994;
Best Local Similarity   21.2%;  Pred. No. 26;
Matches    62;  Conservative    46;  Mismatches   80;  Indels  104;  Gaps   14;

Qy   21 KNEKNRPSLKSLKTDN---RPEKSKCKPLWGKVFYLDLPSVTIS--EKLQKDIKDLGGR 74
Db  356 KSTFSRSLLKTNKNDVLVRPTIPRISP---PLYNLKFSIGLKNQHKELRKRIENLGGK 412

Qy   75 VEEFLSKDISYLISNKKE------AKFAQTLG-RISPV-----------105
Db  413 FEVKISENTIAIISTELETQKKSTRMKFAEELGIHIVPIEFLDFVEADTEGAIKYINSTC 472
```

Db  473 ICSWGTDPKSRIPKETTRSLNSNSIYTRSMPVSRTFVKVDGLAVDPDSGLEDIA--HVY 529
Qy  153 IPSNSILSNALSWGVKILHIDIR-----YY----IEQKKKELYLLRKKSSTSVRDGKRV 203

Qy  204 GSG-------------AQKT-----------RTGRLKKPFVKVED 224
Db  530 VDSNNKYS-------VVLGLTDIQRNKNSYYKVQLLKADKKEKYWIFRSWGRI--GTNI 579
Db  580 GNSKLEEFDTSESAKRNFKEIYADKTGNEYEQRDNFVKRTGRMYPIEIQYDD 631

RESULT 7
C71927
cag island protein - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
C;Variety: strain J99
A;Reference number: A71800; MUID:99120557
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
;Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F
Nature 397, 176-180, 1999
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
A;Accession: C71927
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <ARN>
A;Cross-references: GB1:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD06053.1; PID:g415
A;Experimental source: strain J99
C;Genetics:
A;Gene: orf7

```
Query Match             7.5%;  Score 90;  DB 2;  Length 219;
Best Local Similarity   24.0%;
Matches    46;  Conservative    31;  Mismatches   71;  Indels   44;  Gaps    8;

Qy   10 SKGHFQGFGIQVRNEKNRPSLKSLKTDNRPEKSK-------CKPLWGKVFYLDLPSVT 59
Db   53 NKGDCKGVLREINGSMK--MVCLHCENTPIMEKVESGRGGAYACKNCNRKFYFIDLAKQN 110

Qy   60 ISEK-LQKDIKDLGGRVEEFLSKDISYLI----SNKKEAKFAQTLGRISPVPSPESAYT 113
Db  111 ERKDLEKEKELLNKIERQKIRHLERFILAGVKANIKENSFF----LGCKNYPKCEWT 165

Qy  114 AETTSPHPSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHID 173
Db  166 A-------SMDSQDLKCPKCNRLMKRK------KNFKNNEFFTATSLTLNAIEFCL----- 208

Qy  174 DIRYYIEQKKKE 185
Db  209 ----YINIKKKE 216
```

RESULT 8
S26006
hypothetical protein - liverwort (Marchantia polymorpha) mitochondrion
C;Species: mitochondrion Marchantia polymorpha
C;Date: 07-May-1993
A;Reference number: S26006
R;Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.;
J. Mol. Biol. 223, 1-7, 1992
A;Title: Gene organization deduced from the complete sequence of liverwort Marchantia
A;Accession: S26006
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-710 <ODA>
A;Cross-references: EMBL:M68929; NID:g786182; PIDN:AAC09460.1; PID:g786246
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
C;Genetics:
A;Genome: mitochondrion

Qy  106 -----PSPESAYTAETTSPHPSHD-------GSSFKSPDTVCLSRGKLLVEKAIKDHDF 152

C;Keywords: mitochondrion

RESULT 9
T48502
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48502
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24490
A;Molecule type: DNA
A;Residues: 1-871 <BEV>
A;Cross-references: EMBL:AL163815
A;Experimental source: cultivar Columbia; BAC clone F15N18
A;Genetics:
A;Map position: 5
A;Introns: 562/2; 602/3; 622/3

hypothetical protein F15N18.20 - Arabidopsis thaliana

Query Match     7.4%;  Score 89;  DB 2;  Length 871;
Best Local Similarity  20.2%;  Pred. No. 31;
Matches 37;  Conservative 43;  Mismatches 65;  Indels 38;  Gaps 8;

```
Qy    30  KSLKTDNRPEKSKCKPLWGKVFYLDLP--SVTISEKLQKDIKDLGGRVEEFLSKDISYLI    87
              :::::  ::     :::  :::! :!   ::::::::::::::::::: :::!:!::::
Db    70  KATQTESAPQK-PAKPVVNKKQHVPPQRSVKAMEEVNESVR---SKMRESLASALALVK   125

Qy    88  SNKKEAKFAQTLGRI-SPVPSPESAYTAETTSPH----PSHDGSSFKSPDTVCLSRGK--   140
              :::!::!:: ::: :::!:  :::::::::::        :::  :::::::: :::
Db   126  KDDDSPKGKENIGTVETFVVTIQENTQSFQPASPASISVPVGEGTMSEMPTSVESSVQKDS  185

Qy   141  ----LLVEKAIK--------DHDFIPSNSIL-------SNALSWGVKILHIDDI   175
                  :::::!:        ::::!::::!         ::!:::!::!::!:
Db   186  EIPVDIMMEDVIKFNVLKSQVDEVFPRDNVPFTDIIFPNDDLLHGNELSWDLEVSDLGET  245

Qy   176  RYY  178
          :::
Db   246  KDY  248
```

RESULT 10
B70126
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 11-Jan-2000
C;Accession: B70126
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: B70126
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1119 <KLE>
A;Cross-references: GB:AE001131; GB:AE000783; NID:92688098; PIDN:AAC66595.1; PID:9268
A;Experimental source: strain B31
A;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat
F;742-774/Domain: tetratricopeptide repeat homology #status atypical <TT1>
F;775-808/Domain: tetratricopeptide repeat homology <TT2>
F;809-842/Domain: tetratricopeptide repeat homology <TT3>
F;843-876/Domain: tetratricopeptide repeat homology <TT4>
F;911-943/Domain: tetratricopeptide repeat homology <TT5>
F;944-977/Domain: tetratricopeptide repeat homology <TT6>
F;979-1010/Domain: tetratricopeptide repeat homology #status atypical <TT7>
F;1011-1044/Domain: tetratricopeptide repeat homology <TT8>

surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete

Query Match     7.4%;  Score 89;  DB 2;  Length 1119;
Best Local Similarity  22.8%;  Pred. No. 43;
Matches 58;  Conservative 45;  Mismatches 101;  Indels 50;  Gaps 15;

```
Qy    14  FQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGG    73
              :: !:::!::::::::::::::::: :::! ::!:: ::::::::::::::::::::::
Db    22  NEKNRPSLKSLKTDNRPEKSKCKP--LWGKVF--YLDLPSVTISEKLQKDIKDLG----   72

Qy    74  RVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSS----F   128
          :::::::::::  ::!:::::::::::    :  :::::::!::  :::      :
Db    73  -GRVEEFLSKDIS--------YLISNKKEAKFAQTL-----GRISP--VPSPESAY---  112

Qy   129  KSPDTVCLSRGKLLVEKAIK-----------DHDFIPSNSILS--   173
          :::::::::::::::::! !            :::::::::!::
Db   113  ---TAETTSPHPSHD-GSSFKSPDTVCLSRGKLLVEKAIKD-HDFI---PSNSILSNA  162

Qy   174  DIRYYIEQKKELYLLKKSSTSVR   197
          ::!::!::: !:!::!:!:!:::
Db   163  LSWGVKILHIDDIRYYIEQ--KKRELYLLKKSSTSVRDGG---KRVGSGAQKTRTGRLKKP  218

Qy   226  FPGSI--AKSAKATTKTKANKGESRP-----VTPPALGRVFYEDIYNI--DNLRAGYKRLKG  278
              :::   :::  : :: :::::!       :! ::: ::  :::!    ::!::!:!:::
Db   219  FVKVEDMSQSPAVH  232

Db   365  EIFQLDKEDKKPQYLEDLKSKVHSIKPIDLENTKSR-QQAIKDLNEFLKNPNDAQASKT  423

Qy   279  NVAPGIDGRTKADMTDKALEKLSKELRQAYAPKP----AKRIIITKP----DGGSRPLSIA  332

Qy   333  STVDKVVQSTLKELVEPHFESLFRDSSHGFRPGRSCHKALRDLRYSWTALTWLVQI----  388
Db   424  LAQANKIQHEEDLKSKVHSIKPIDLENTKSRQQAIKDLNEFLKNPNDAQASKTLAQANK  483

Qy   389  DIKKDFDKIHIHDL-LIKEMESVLR  411
          :: ::: ::!::: :::!!:::!::
Db   484  IQHLEDLKSK--VH  495
```

RESULT 11
B70242
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: B70242
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: B70242
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-392 <KLE>
A;Cross-references: GB:AE000789; NID:92690079; PIDN:AAC66184.1; PID:92690081; TIGR:BB
A;Experimental source: strain B31
A;Genetics:
A;Genome: plasmid
A;Superfamily: Lyme disease spirochete plasmid conserved hypothetical protein BBG06

conserved hypothetical protein BBI19 - Lyme disease spirochete plasmid I/lp28-4

Query Match     7.3%;  Score 87.5;  DB 2;  Length 392;

Best Local Similarity 22.5%; Conservative 32%; Mismatches 15;
Matches 41; Pred. No, 78; Indels 31; Gaps 5;

```
Qy   22 NEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYL----DLPSVTISEKLQKDIKDLGGRVE 76
           |:|:| ||:||:|                     ||:|||             ::|
Db  168 NNNKEEEEDIRNNKIEKCCIKKYFNKCNFLSEEAKSILELNISRKNTIEIKIKKIE 227

Qy   77 EFLSKDISYLISNKKEAKFAQTLGRISPVPSP------ESAYTAETTSPHPSHD 124
           :||                        | ||             : ||
Db  228 TDLTKNKNKYCGFKKKQKMLKEILSKTKKQLEKKGYDTKQLKLKIENIYKSYKTKPHFIIE 287

Qy  125 GSSFKSPDTVCLSRGKLLVEKAIKIDHDFIPSNSILSNALSWGVKILHIDDIRYYIEQKKK 184
           :                                                      :
Db  288 NKKYKDLKI--------RLKLEKSIE---IKKESIPKKYIHKVNIFNI----LIEQLKK 333

Qy  185 EL 186
        ||
Db  334 EL 335
```

RESULT 12
T28810
hypothetical protein C54D1.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T28810
R;Minx, M.
submitted to the EMBL Data Library, January 1996
A;Description: The sequence of C. elegans cosmid C54D1.
A;Reference number: Z20527
A;Accession: T28810
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-506 <MIN>
A;Cross-references: EMBL:U46673; PIDN:AAC48151.1; GSPDB:GN00028; CESP:C54D1.4
A;Experimental source: strain Bristol N2; clone C54D1
A;Genetics:
C;Gene: CESP:C54D1.4
A;Map position: X
A;Introns: 78/2; 121/2; 180/3; 296/2; 332/3; 384/3; 421/3; 464/3
A;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 24.0%; Score 87.5; DB 2; Length 506;
Best Local Similarity 7.3%; Pred. No. 21;
Matches 41; Conservative 29; Mismatches 52; Indels 49; Gaps 8;

```
Qy   50 VFYLDLP------SVTISEKLQKDIKDLGGRVEEEFLSKDISYLISNKKEAKFAQTLGRIS 103
           :|:|:|       ||:|                   :|:                |
Db  239 VDHDVPLISPTGSTVIGKKIQED----GAK---LNKKVSLEMGGK 277

Qy  104 PVPSPESAYTAETTSPHPSHDGSSFKSPDTVCLSRGKLLVEKAIKIDHDFI-------- 153
Db  278 --NPGIVYSNYRKSDIASIARSSFLNQGEICLCTSRLFVQKPI-PADFVKSYVEEAKKF 333

Qy  154 ----PSNILSNALSWGVKILHIDDIRYYIEQKKKE-LYLLKKSSTSVRDG 199
            :|:|  | ||              :|||      :: |||   |:||:
Db  334 TVGDPTTQVQIGAMNSKV----HYEKVKSYIELAKKEGADILCGGVTTIONG 381
```

RESULT 13
A71564
hypothetical protein CT051 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 15-Sep-2000
C;Accession: A71564
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A;Reference number: A71570; MUID:99000809
A;Accession: A71564
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-520 <ARN>
A;Cross-references: GB:AE001279; GB:AE001273; NID:g3328434; PIDN:AAC67642.1; PID:g332
A;Genetics:
A;Gene: CT051
C;Superfamily: Chlamydia trachomatis hypothetical protein CT051

Query Match 21.9%; Score 87.5; DB 2; Length 520;
Best Local Similarity 7.3%; Pred. No. 21;
Matches 65; Conservative 39; Mismatches 94; Indels 99; Gaps 15;

```
Qy    2 NSGAMRIHSKGHFQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVT-- 59
Db   76 NPGYLPIHG-GNMTGNIINMGTH---SVFNLK---QPEKPK------IELPSEIDK 117

Qy   60 ------------ISEKLQKDIKDLGGRVEEFLSKDISY------------- 85
                        :|||:    ||::|| |:|  |:|
Db  118 PKDPREEDGFAEKTAEEQEOEIKEYNTKLAEYQKKIDDYNAWWEAFYSEAATVKVKGIV 177

Qy  127 SFKSPDTVCLSRGKLLVEKAIKIDHDFIPSNS------------ILSNALSWGVKILHIDIRY- 177
           |:                                         |||||||||| ||||||
Db  178 DKILNNDKLSTALNSATEVEKKIALAQKALGIEITINPDADINPDTDQETPDAPVADTE 237

Qy  178 ----YIEQKKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVVEDMSQSPA 230
            ::|   |:| |:|  | :|| |:||        :| :| ||:|
Db  238 EKESPP---LSYNDL--PSVIKNSQFVVTQSQNKITGDLKMTNAQIANIKTPDTGDSNYA 292

Qy  293 ANVTYLESKLKQPQRAFLSNTLPTE-----SSSSISLNGHI-PWLSTTNGSSSPA 341
```

RESULT 14
B69783
transporter homolog ydgH - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: B69783
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fujita, M.; Fuma, S.; Galizzi, A.; Gal
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: B69783
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-885 <KUN>
A;Cross-references: GB:Z99106; GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12384.1;
A;Experimental source: strain 168
A;Genetics:
A;Gene: ydgH

Query Match 23.3%; Score 87.5; DB 2; Length 885;
Best Local Similarity 7.3%; Pred. No. 42;
Matches 47; Conservative 30; Mismatches 74; Indels 51; Gaps 8;

```
Qy    9 HSKGHFQGGIQVKNEKNRPSLKSLKTDN-------------RPEKSKCKPLWGKVFYLD 54
           :|||   | | |  |:      | :|              ||||||||||||||||
Db   34 NQKGQAQLPADAVSERANAILKQAGEDNNSISVVFTLDNAIKKETENQLRIIIDKIKKD 93
```

```
Qy      55  -LPSVTISEKLQKDIKDLGGRVEEFLSKD---------ISYLISNKEAKFAQTLGRISPVP 106
                :  ||  |          :|||  |        :||| ||    :|   ||
Db      94  GVEEVTSPLSAEKEVKD-----QLMSKDKKTVLMPVVITGSDKRAEKIADEIYQI--VP 145

Qy     107  SPESAYTAETT------SPHPSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIP------ 154
                :||             | ||  ||:||| |  |||
Db     146  DDLTAYITGASLINQDFAHSSEEGLKKTEVITVCLIIGLLLIVFRSVVTPFIPIVVVGFS 205

Qy     155  ---SNSILSNALSWGVKILHID 173
               :|||              :|:|
Db     206  YLISQSIL------GILVYNVD 221
```

RESULT  15

F82315
conserved hypothetical protein VC0503 [imported] - Vibrio cholerae (strain N16961 serogr
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: F82315
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82315
A;Accession: F82315
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-426 <HEI>
A;Cross-references: GB:AE004136; GB:AE003852; NID:g9654921; PIDN:AAF93673.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0503
A;Map position: 1

Query Match             7.2%;  Score 87;  DB 2;  Length 426;
Best Local Similarity   24.1%;  Pred. No. 18;
Matches   46;  Conservative   28;  Mismatches   67;  Indels   50;  Gaps   9;

```
Qy      60  ISEKLQKDIKDLGGRVEEFLSKDISYL--ISNKKEAKFAQTLGRISPVPSPESAYTAETT 117
                |||| |  |  ||| |  | |:| :|   :|| |:||| ::  |  :     : :
Db     229  ISFKLAK------GDVSAFLAEDGRFYDRAGNSLERAFNR----YPV---DKAYRQITS 274

Qy     118  SPHPS---------HDGSSFKSP---DTVCLSRGKLLVEKAIKDHDFIPSNSILSNAL 163
                 :|                |  ||:|:|:|::|   |:|::  :
Db     275  GFNPKRKHPVTGRVVPHNGTDFATPIGAPVYSTGDGKIV---VRKHPYAGNYLVIEHNS 331

Qy     164  SWGVKILHIDDIRYYIEQKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVE 223
                 ||:| |  ||    || ::  :|     :| |     |  |  | ::| :
Db     332  VYKTRYLHLDKI-----------------LVKKGQLVKRQQKIALAGA----TGRLTGPHLHFE 374

Qy     224  DMSQSPAVHLM 234
                 : |:
Db     375  VLVRNRPVDAM 385
```

Search completed: December 27, 2001, 16:54:46
Job time: 276 sec